autogluon_all

October 7, 2023

[11]:
```
import pandas as pd
import numpy as np



import warnings
warnings.filterwarnings("ignore")

def fix_datetime(X, name):
    # Convert 'date_forecast' to datetime format and replace original column␣
 ↪with 'ds'
    X['ds'] = pd.to_datetime(X['date_forecast'])
    X.drop(columns=['date_forecast'], inplace=True, errors='ignore')
    X.sort_values(by='ds', inplace=True)
    X.set_index('ds', inplace=True)

    # Drop rows where the minute part of the time is not 0
    X = X[X.index.minute == 0]
    return X



def convert_to_datetime(X_train_observed, X_train_estimated, X_test, y_train):
    X_train_observed = fix_datetime(X_train_observed, "X_train_observed")
    X_train_estimated = fix_datetime(X_train_estimated, "X_train_estimated")
    X_test = fix_datetime(X_test, "X_test")


    X_train_observed["estimated_diff_hours"] = 0
    X_train_observed["is_estimated"] = False
    X_train_estimated["estimated_diff_hours"] = (X_train_estimated.index - pd.
 ↪to_datetime(X_train_estimated["date_calc"])).dt.total_seconds() / 3600
    X_test["estimated_diff_hours"] = (X_test.index - pd.
 ↪to_datetime(X_test["date_calc"])).dt.total_seconds() / 3600

    X_train_estimated["is_estimated"] = True
    X_test["is_estimated"] = True
```

```
    X_train_estimated["estimated_diff_hours"] =␣
 ↪X_train_estimated["estimated_diff_hours"].astype('int64')
    # the filled once will get dropped later anyways, when we drop y nans
    X_test["estimated_diff_hours"] = X_test["estimated_diff_hours"].fillna(-50).
 ↪astype('int64')


    X_train_estimated.drop(columns=['date_calc'], inplace=True)
    X_test.drop(columns=['date_calc'], inplace=True)

    y_train['ds'] = pd.to_datetime(y_train['time'])
    y_train.drop(columns=['time'], inplace=True)
    y_train.sort_values(by='ds', inplace=True)
    y_train.set_index('ds', inplace=True)

    return X_train_observed, X_train_estimated, X_test, y_train




def preprocess_data(X_train_observed, X_train_estimated, X_test, y_train,␣
 ↪location):
    # convert to datetime
    X_train_observed, X_train_estimated, X_test, y_train =␣
 ↪convert_to_datetime(X_train_observed, X_train_estimated, X_test, y_train)

    y_train["y"] = y_train["pv_measurement"].astype('float64')
    y_train.drop(columns=['pv_measurement'], inplace=True)

    X_train = pd.concat([X_train_observed, X_train_estimated,␣
 ↪X_train_estimated, X_train_estimated, X_train_estimated, X_train_estimated],␣
 ↪axis=0)
    # weight the estimated X_train higher



    # clip all y values to 0 if negative
    y_train["y"] = y_train["y"].clip(lower=0)

    X_train = pd.merge(X_train, y_train, how="outer", left_index=True,␣
 ↪right_index=True)

    X_train["location"] = location
    X_test["location"] = location
```

```
    return X_train, X_test
# Define locations
locations = ['A', 'B', 'C']

X_trains = []
X_tests = []
# Loop through locations
for loc in locations:
    print(f"Processing location {loc}...")
    # Read target training data
    y_train = pd.read_parquet(f'{loc}/train_targets.parquet')

    # Read estimated training data and add location feature
    X_train_estimated = pd.read_parquet(f'{loc}/X_train_estimated.parquet')

    # Read observed training data and add location feature
    X_train_observed= pd.read_parquet(f'{loc}/X_train_observed.parquet')

    # Read estimated test data and add location feature
    X_test_estimated = pd.read_parquet(f'{loc}/X_test_estimated.parquet')

    # Preprocess data
    X_train, X_test = preprocess_data(X_train_observed, X_train_estimated,␣
 ↪X_test_estimated, y_train, loc)

    X_trains.append(X_train)
    X_tests.append(X_test)

# Concatenate all data and save to csv
X_train = pd.concat(X_trains)
X_test = pd.concat(X_tests)
```

```
Processing location A…
Processing location B…
Processing location C…
```

1 Feature enginering

[12]:
```
# temporary
X_train["hour"] = X_train.index.hour
X_train["weekday"] = X_train.index.weekday
X_train["month"] = X_train.index.month
X_train["year"] = X_train.index.year

X_test["hour"] = X_test.index.hour
X_test["weekday"] = X_test.index.weekday
X_test["month"] = X_test.index.month
```

```
X_test["year"] = X_test.index.year


to_drop = ["snow_drift:idx", "snow_density:kgm3"]

X_train.drop(columns=to_drop, inplace=True)
X_test.drop(columns=to_drop, inplace=True)

X_train.dropna(subset=['y'], inplace=True)
X_train.to_csv('X_train_raw.csv', index=True)
X_test.to_csv('X_test_raw.csv', index=True)
```

[13]:
```
import autogluon.eda.auto as auto
auto.dataset_overview(train_data=X_train, test_data=X_test, label="y",␣
↪sample=None)
```

train_data dataset summary

	count	unique	top	freq	mean	\
absolute_humidity_2m:gm3	136651	165			5.463406	
air_density_2m:kgm3	136651	293			1.262643	
ceiling_height_agl:m	104000	40993			2936.124268	
clear_sky_energy_1h:J	136651	48602			435987.8125	
clear_sky_rad:W	136651	7815			121.115158	
cloud_base_agl:m	121212	34862			1666.626587	
dew_or_rime:idx	136651	3			0.00431	
dew_point_2m:K	136651	436			273.853149	
diffuse_rad:W	136651	2870			34.930195	
diffuse_rad_1h:J	136651	48553			125727.84375	
direct_rad:W	136651	5296			43.46027	
direct_rad_1h:J	136651	41885			156453.46875	
effective_cloud_cover:p	136651	1001			67.058701	
elevation:m	136651	3			11.323496	
estimated_diff_hours	136651	26			10.691213	
fresh_snow_12h:cm	136651	125			0.17007	
fresh_snow_1h:cm	136651	39			0.014023	
fresh_snow_24h:cm	136651	161			0.331657	
fresh_snow_3h:cm	136651	70			0.042357	
fresh_snow_6h:cm	136651	96			0.08499	
hour	136724	24			11.502063	
is_day:idx	136651	2			0.447322	
is_estimated	136651	2	False	82026		
is_in_shadow:idx	136651	2			0.603062	
location	136724	3	A	51661		
month	136724	12			5.950192	
msl_pressure:hPa	136651	874			1009.870605	
precip_5min:mm	136651	64			0.005298	
precip_type_5min:idx	136651	7			0.08169	

pressure_100m:hPa	136651	888	996.142151
pressure_50m:hPa	136651	897	1002.305237
prob_rime:p	136651	700	0.869597
rain_water:kgm2	136651	11	0.00773
relative_humidity_1000hPa:p	136651	788	73.116318
sfc_pressure:hPa	136651	902	1008.496094
snow_depth:cm	136651	165	0.170452
snow_melt_10min:mm	136651	19	0.000336
snow_water:kgm2	136651	42	0.087429
sun_azimuth:d	136651	69692	182.439377
sun_elevation:d	136651	49376	-3.786569
super_cooled_liquid_water:kgm2	136651	15	0.050161
t_1000hPa:K	136651	447	278.169525
total_cloud_cover:p	136651	1001	73.306854
visibility:m	136651	85686	33385.910156
weekday	136724	7	3.011176
wind_speed_10m:ms	136651	119	3.067309
wind_speed_u_10m:ms	136651	188	0.455725
wind_speed_v_10m:ms	136651	167	0.761585
wind_speed_w_1000hPa:ms	136651	3	-0.000084
y	136724	12430	247.857695
year	136724	6	2021.333541

	std	min	25%	\
absolute_humidity_2m:gm3	2.526241	0.5	3.7	
air_density_2m:kgm3	0.035762	1.139	1.238	
ceiling_height_agl:m	2539.34668	27.799999	1061.400024	
clear_sky_energy_1h:J	753586.4375	0.0	0.0	
clear_sky_rad:W	209.874939	0.0	0.0	
cloud_base_agl:m	1796.03125	27.4	539.900024	
dew_or_rime:idx	0.253486	-1.0	0.0	
dew_point_2m:K	6.624437	247.300003	269.399994	
diffuse_rad:W	57.239651	0.0	0.0	
diffuse_rad_1h:J	203732.015625	0.0	0.0	
direct_rad:W	105.165146	0.0	0.0	
direct_rad_1h:J	374123.8125	0.0	0.0	
effective_cloud_cover:p	35.726498	0.0	40.099998	
elevation:m	7.849141	6.0	6.0	
estimated_diff_hours	13.81461	0.0	0.0	
fresh_snow_12h:cm	0.861506	0.0	0.0	
fresh_snow_1h:cm	0.123339	0.0	0.0	
fresh_snow_24h:cm	1.356192	0.0	0.0	
fresh_snow_3h:cm	0.306691	0.0	0.0	
fresh_snow_6h:cm	0.526908	0.0	0.0	
hour	6.919173	0.0	6.0	
is_day:idx	0.497219	0.0	0.0	
is_estimated				
is_in_shadow:idx	0.489265	0.0	0.0	

location				
month	3.798353	1.0	3.0	
msl_pressure:hPa	13.545425	944.299988	1001.099976	
precip_5min:mm	0.03049	0.0	0.0	
precip_type_5min:idx	0.387124	0.0	0.0	
pressure_100m:hPa	13.422117	929.799988	987.5	
pressure_50m:hPa	13.488344	935.599976	993.599976	
prob_rime:p	5.886336	0.0	0.0	
rain_water:kgm2	0.038118	0.0	0.0	
relative_humidity_1000hPa:p	14.40036	19.5	63.299999	
sfc_pressure:hPa	13.556752	941.400024	999.700012	
snow_depth:cm	1.085614	0.0	0.0	
snow_melt_10min:mm	0.004806	-0.0	0.0	
snow_water:kgm2	0.237246	0.0	0.0	
sun_azimuth:d	102.127739	0.008	94.625999	
sun_elevation:d	23.617214	-49.979	-21.3685	
super_cooled_liquid_water:kgm2	0.105082	0.0	0.0	
t_1000hPa:K	6.182547	257.899994	274.0	
total_cloud_cover:p	35.852039	0.0	49.5	
visibility:m	18111.511719	130.600006	17141.099609	
weekday	2.006202	0.0	1.0	
wind_speed_10m:ms	1.782652	0.0	1.7	
wind_speed_u_10m:ms	2.880435	-7.3	-1.6	
wind_speed_v_10m:ms	1.871658	-9.3	-0.5	
wind_speed_w_1000hPa:ms	0.006292	-0.1	-0.0	
y	717.454239	-0.0	0.0	
year	1.376591	2018.0	2020.0	

	50%	75%	max	\
absolute_humidity_2m:gm3	4.9	6.9	17.5	
air_density_2m:kgm3	1.263	1.285	1.441	
ceiling_height_agl:m	1974.099976	4185.275391	12431.299805	
clear_sky_energy_1h:J	0.0	552515.3125	3006697.25	
clear_sky_rad:W	0.0	152.199997	835.299988	
cloud_base_agl:m	1065.0	1980.5	11688.900391	
dew_or_rime:idx	0.0	0.0	1.0	
dew_point_2m:K	273.5	278.5	293.799988	
diffuse_rad:W	0.0	55.5	340.100006	
diffuse_rad_1h:J	0.0	198568.3125	1182265.375	
direct_rad:W	0.0	18.299999	684.299988	
direct_rad_1h:J	0.0	70565.648438	2445897.0	
effective_cloud_cover:p	81.800003	99.599998	100.0	
elevation:m	7.0	24.0	24.0	
estimated_diff_hours	0.0	24.0	39.0	
fresh_snow_12h:cm	0.0	0.0	37.400002	
fresh_snow_1h:cm	0.0	0.0	7.1	
fresh_snow_24h:cm	0.0	0.0	37.400002	
fresh_snow_3h:cm	0.0	0.0	20.6	

fresh_snow_6h:cm	0.0	0.0	34.0
hour	12.0	17.0	23.0
is_day:idx	0.0	1.0	1.0
is_estimated			
is_in_shadow:idx	1.0	1.0	1.0
location			
month	5.0	10.0	12.0
msl_pressure:hPa	1010.400024	1019.200012	1044.099976
precip_5min:mm	0.0	0.0	1.38
precip_type_5min:idx	0.0	0.0	6.0
pressure_100m:hPa	996.799988	1005.5	1030.900024
pressure_50m:hPa	1003.0	1011.700012	1037.300049
prob_rime:p	0.0	0.0	97.199997
rain_water:kgm2	0.0	0.0	1.4
relative_humidity_1000hPa:p	75.300003	84.5	100.0
sfc_pressure:hPa	1009.099976	1017.900024	1043.800049
snow_depth:cm	0.0	0.0	18.299999
snow_melt_10min:mm	0.0	-0.0	0.18
snow_water:kgm2	0.0	0.1	6.9
sun_azimuth:d	179.524994	270.320496	359.997009
sun_elevation:d	-2.985	11.878	49.917999
super_cooled_liquid_water:kgm2	0.0	0.1	1.4
t_1000hPa:K	277.299988	282.100006	303.299988
total_cloud_cover:p	96.0	100.0	100.0
visibility:m	36908.898438	48746.648438	76737.796875
weekday	3.0	5.0	6.0
wind_speed_10m:ms	2.7	4.1	15.2
wind_speed_u_10m:ms	0.1	2.3	12.2
wind_speed_v_10m:ms	0.8	1.9	9.0
wind_speed_w_1000hPa:ms	0.0	-0.0	0.1
y	0.0	113.85	5733.42
year	2021.0	2023.0	2023.0

	dtypes	missing_count	missing_ratio	raw_type	\
absolute_humidity_2m:gm3	float32	73	0.000534	float	
air_density_2m:kgm3	float32	73	0.000534	float	
ceiling_height_agl:m	float32	32724	0.239343	float	
clear_sky_energy_1h:J	float32	73	0.000534	float	
clear_sky_rad:W	float32	73	0.000534	float	
cloud_base_agl:m	float32	15512	0.113455	float	
dew_or_rime:idx	float32	73	0.000534	float	
dew_point_2m:K	float32	73	0.000534	float	
diffuse_rad:W	float32	73	0.000534	float	
diffuse_rad_1h:J	float32	73	0.000534	float	
direct_rad:W	float32	73	0.000534	float	
direct_rad_1h:J	float32	73	0.000534	float	
effective_cloud_cover:p	float32	73	0.000534	float	
elevation:m	float32	73	0.000534	float	

estimated_diff_hours	float64	73	0.000534	float
fresh_snow_12h:cm	float32	73	0.000534	float
fresh_snow_1h:cm	float32	73	0.000534	float
fresh_snow_24h:cm	float32	73	0.000534	float
fresh_snow_3h:cm	float32	73	0.000534	float
fresh_snow_6h:cm	float32	73	0.000534	float
hour	int64			int
is_day:idx	float32	73	0.000534	float
is_estimated	object	73	0.000534	object
is_in_shadow:idx	float32	73	0.000534	float
location	object			object
month	int64			int
msl_pressure:hPa	float32	73	0.000534	float
precip_5min:mm	float32	73	0.000534	float
precip_type_5min:idx	float32	73	0.000534	float
pressure_100m:hPa	float32	73	0.000534	float
pressure_50m:hPa	float32	73	0.000534	float
prob_rime:p	float32	73	0.000534	float
rain_water:kgm2	float32	73	0.000534	float
relative_humidity_1000hPa:p	float32	73	0.000534	float
sfc_pressure:hPa	float32	73	0.000534	float
snow_depth:cm	float32	73	0.000534	float
snow_melt_10min:mm	float32	73	0.000534	float
snow_water:kgm2	float32	73	0.000534	float
sun_azimuth:d	float32	73	0.000534	float
sun_elevation:d	float32	73	0.000534	float
super_cooled_liquid_water:kgm2	float32	73	0.000534	float
t_1000hPa:K	float32	73	0.000534	float
total_cloud_cover:p	float32	73	0.000534	float
visibility:m	float32	73	0.000534	float
weekday	int64			int
wind_speed_10m:ms	float32	73	0.000534	float
wind_speed_u_10m:ms	float32	73	0.000534	float
wind_speed_v_10m:ms	float32	73	0.000534	float
wind_speed_w_1000hPa:ms	float32	73	0.000534	float
y	float64			float
year	int64			int

	variable_type	special_types
absolute_humidity_2m:gm3	numeric	
air_density_2m:kgm3	numeric	
ceiling_height_agl:m	numeric	
clear_sky_energy_1h:J	numeric	
clear_sky_rad:W	numeric	
cloud_base_agl:m	numeric	
dew_or_rime:idx	category	
dew_point_2m:K	numeric	
diffuse_rad:W	numeric	

diffuse_rad_1h:J	numeric
direct_rad:W	numeric
direct_rad_1h:J	numeric
effective_cloud_cover:p	numeric
elevation:m	category
estimated_diff_hours	numeric
fresh_snow_12h:cm	numeric
fresh_snow_1h:cm	numeric
fresh_snow_24h:cm	numeric
fresh_snow_3h:cm	numeric
fresh_snow_6h:cm	numeric
hour	numeric
is_day:idx	category
is_estimated	category
is_in_shadow:idx	category
location	category
month	category
msl_pressure:hPa	numeric
precip_5min:mm	numeric
precip_type_5min:idx	category
pressure_100m:hPa	numeric
pressure_50m:hPa	numeric
prob_rime:p	numeric
rain_water:kgm2	category
relative_humidity_1000hPa:p	numeric
sfc_pressure:hPa	numeric
snow_depth:cm	numeric
snow_melt_10min:mm	category
snow_water:kgm2	numeric
sun_azimuth:d	numeric
sun_elevation:d	numeric
super_cooled_liquid_water:kgm2	category
t_1000hPa:K	numeric
total_cloud_cover:p	numeric
visibility:m	numeric
weekday	category
wind_speed_10m:ms	numeric
wind_speed_u_10m:ms	numeric
wind_speed_v_10m:ms	numeric
wind_speed_w_1000hPa:ms	category
y	numeric
year	category

test_data dataset summary

	count	unique	top	freq	mean	\
absolute_humidity_2m:gm3	2160	106			8.206482	
air_density_2m:kgm3	2160	153			1.232807	
ceiling_height_agl:m	1473	1391			2938.389648	

clear_sky_energy_1h:J	2160	1807			1227746.75
clear_sky_rad:W	2160	1044			341.056641
cloud_base_agl:m	1879	1771			1797.160156
dew_or_rime:idx	2160	3			0.040741
dew_point_2m:K	2160	202			280.783203
diffuse_rad:W	2160	985			84.915688
diffuse_rad_1h:J	2160	1806			305696.5
direct_rad:W	2160	916			114.279816
direct_rad_1h:J	2160	1634			411408.875
effective_cloud_cover:p	2160	590			64.113792
elevation:m	2160	3			12.333333
estimated_diff_hours	2160	24			27.5
fresh_snow_12h:cm	2160	2			0.000185
fresh_snow_1h:cm	2160	2			0.000185
fresh_snow_24h:cm	2160	2			0.000185
fresh_snow_3h:cm	2160	2			0.000185
fresh_snow_6h:cm	2160	2			0.000185
hour	2160	24			11.5
is_day:idx	2160	2			0.795833
is_estimated	2160	1	True	2160	
is_in_shadow:idx	2160	2			0.24537
location	2160	3	A	720	
month	2160	3			5.666667
msl_pressure:hPa	2160	321			1016.805786
precip_5min:mm	2160	27			0.00775
precip_type_5min:idx	2160	3			0.065741
pressure_100m:hPa	2160	359			1002.970825
pressure_50m:hPa	2160	356			1009.007202
prob_rime:p	2160	3			0.01588
rain_water:kgm2	2160	8			0.013056
relative_humidity_1000hPa:p	2160	538			70.920792
sfc_pressure:hPa	2160	363			1015.070374
snow_depth:cm	2160	1			0.0
snow_melt_10min:mm	2160	1			0.0
snow_water:kgm2	2160	16			0.060972
sun_azimuth:d	2160	1830			183.166199
sun_elevation:d	2160	1623			20.292332
super_cooled_liquid_water:kgm2	2160	7			0.065463
t_1000hPa:K	2160	254			284.737732
total_cloud_cover:p	2160	553			69.298981
visibility:m	2160	2155			33304.636719
weekday	2160	7			3.233333
wind_speed_10m:ms	2160	83			2.946759
wind_speed_u_10m:ms	2160	123			1.650694
wind_speed_v_10m:ms	2160	80			-0.187176
wind_speed_w_1000hPa:ms	2160	2			0.000324
year	2160	1			2023.0

	std	min	25%	\
absolute_humidity_2m:gm3	2.201396	3.2	6.6	
air_density_2m:kgm3	0.032116	1.142	1.209	
ceiling_height_agl:m	2913.641113	30.6	891.799988	
clear_sky_energy_1h:J	1104468.625	0.0	64338.124023	
clear_sky_rad:W	307.729095	0.0	13.65	
cloud_base_agl:m	2046.394409	29.799999	486.899994	
dew_or_rime:idx	0.202365	-1.0	0.0	
dew_point_2m:K	4.378817	268.0	277.899994	
diffuse_rad:W	78.422508	0.0	6.925	
diffuse_rad_1h:J	278146.25	0.0	36756.901367	
direct_rad:W	171.838226	0.0	0.0	
direct_rad_1h:J	611480.125	0.0	86.575001	
effective_cloud_cover:p	37.947498	0.0	30.700001	
elevation:m	8.261587	6.0	6.0	
estimated_diff_hours	6.923789	16.0	21.75	
fresh_snow_12h:cm	0.008607	0.0	0.0	
fresh_snow_1h:cm	0.008607	0.0	0.0	
fresh_snow_24h:cm	0.008607	0.0	0.0	
fresh_snow_3h:cm	0.008607	0.0	0.0	
fresh_snow_6h:cm	0.008607	0.0	0.0	
hour	6.923789	0.0	5.75	
is_day:idx	0.403185	0.0	1.0	
is_estimated				
is_in_shadow:idx	0.430406	0.0	0.0	
location				
month	0.596423	5.0	5.0	
msl_pressure:hPa	9.728754	986.099976	1011.5	
precip_5min:mm	0.033776	0.0	0.0	
precip_type_5min:idx	0.249747	0.0	0.0	
pressure_100m:hPa	9.644145	971.799988	997.799988	
pressure_50m:hPa	9.74076	977.700012	1003.799988	
prob_rime:p	0.551282	0.0	0.0	
rain_water:kgm2	0.055256	0.0	0.0	
relative_humidity_1000hPa:p	15.725973	23.9	60.275	
sfc_pressure:hPa	9.840412	983.5	1009.799988	
snow_depth:cm	0.0	0.0	0.0	
snow_melt_10min:mm	0.0	-0.0	-0.0	
snow_water:kgm2	0.219562	0.0	0.0	
sun_azimuth:d	109.193207	8.27	85.359253	
sun_elevation:d	18.681047	-11.617	1.96475	
super_cooled_liquid_water:kgm2	0.115824	0.0	0.0	
t_1000hPa:K	5.839595	273.700012	279.799988	
total_cloud_cover:p	38.41222	0.0	32.799999	
visibility:m	15624.633789	874.400024	19635.100098	
weekday	2.186573	0.0	1.0	
wind_speed_10m:ms	1.733865	0.0	1.5	
wind_speed_u_10m:ms	2.578466	-4.3	-0.2	

wind_speed_v_10m:ms	1.50826	-4.4	-1.3	
wind_speed_w_1000hPa:ms	0.005685	-0.0	0.0	
year	0.0	2023.0	2023.0	

	50%	75%	max	\
absolute_humidity_2m:gm3	8.0	10.0	14.2	
air_density_2m:kgm3	1.238	1.26	1.301	
ceiling_height_agl:m	1553.900024	4021.300049	11468.0	
clear_sky_energy_1h:J	1056303.125	2372037.5	3005707.0	
clear_sky_rad:W	273.849991	646.874985	835.099976	
cloud_base_agl:m	997.799988	2298.300049	11467.799805	
dew_or_rime:idx	0.0	0.0	1.0	
dew_point_2m:K	281.0	284.299988	290.200012	
diffuse_rad:W	73.700001	135.600006	312.600006	
diffuse_rad_1h:J	272526.046875	488256.03125	1086246.25	
direct_rad:W	16.200001	180.399994	668.0	
direct_rad_1h:J	60416.199219	686746.859375	2403444.25	
effective_cloud_cover:p	77.75	100.0	100.0	
elevation:m	7.0	24.0	24.0	
estimated_diff_hours	27.5	33.25	39.0	
fresh_snow_12h:cm	0.0	0.0	0.4	
fresh_snow_1h:cm	0.0	0.0	0.4	
fresh_snow_24h:cm	0.0	0.0	0.4	
fresh_snow_3h:cm	0.0	0.0	0.4	
fresh_snow_6h:cm	0.0	0.0	0.4	
hour	11.5	17.25	23.0	
is_day:idx	1.0	1.0	1.0	
is_estimated				
is_in_shadow:idx	0.0	0.0	1.0	
location				
month	6.0	6.0	7.0	
msl_pressure:hPa	1020.599976	1023.799988	1029.599976	
precip_5min:mm	0.0	0.0	0.34	
precip_type_5min:idx	0.0	0.0	2.0	
pressure_100m:hPa	1006.25	1010.099976	1016.400024	
pressure_50m:hPa	1012.299988	1016.200012	1022.5	
prob_rime:p	0.0	0.0	23.0	
rain_water:kgm2	0.0	0.0	0.7	
relative_humidity_1000hPa:p	73.900002	83.699997	98.900002	
sfc_pressure:hPa	1018.299988	1022.299988	1028.699951	
snow_depth:cm	0.0	0.0	0.0	
snow_melt_10min:mm	0.0	0.0	0.0	
snow_water:kgm2	0.0	0.0	3.4	
sun_azimuth:d	184.236	279.576248	356.984009	
sun_elevation:d	18.54	38.102499	49.902	
super_cooled_liquid_water:kgm2	0.0	0.1	0.6	
t_1000hPa:K	284.799988	288.299988	302.200012	
total_cloud_cover:p	95.300003	100.0	100.0	

visibility:m	37623.050781	45378.099609	63863.800781
weekday	3.0	5.0	6.0
wind_speed_10m:ms	2.7	4.0	8.8
wind_speed_u_10m:ms	1.6	3.525	8.8
wind_speed_v_10m:ms	-0.3	0.8	4.0
wind_speed_w_1000hPa:ms	0.0	0.0	0.1
year	2023.0	2023.0	2023.0

	dtypes	missing_count	missing_ratio	raw_type	\
absolute_humidity_2m:gm3	float32			float	
air_density_2m:kgm3	float32			float	
ceiling_height_agl:m	float32	687	0.318056	float	
clear_sky_energy_1h:J	float32			float	
clear_sky_rad:W	float32			float	
cloud_base_agl:m	float32	281	0.130093	float	
dew_or_rime:idx	float32			float	
dew_point_2m:K	float32			float	
diffuse_rad:W	float32			float	
diffuse_rad_1h:J	float32			float	
direct_rad:W	float32			float	
direct_rad_1h:J	float32			float	
effective_cloud_cover:p	float32			float	
elevation:m	float32			float	
estimated_diff_hours	int64			int	
fresh_snow_12h:cm	float32			float	
fresh_snow_1h:cm	float32			float	
fresh_snow_24h:cm	float32			float	
fresh_snow_3h:cm	float32			float	
fresh_snow_6h:cm	float32			float	
hour	int64			int	
is_day:idx	float32			float	
is_estimated	bool			bool	
is_in_shadow:idx	float32			float	
location	object			object	
month	int64			int	
msl_pressure:hPa	float32			float	
precip_5min:mm	float32			float	
precip_type_5min:idx	float32			float	
pressure_100m:hPa	float32			float	
pressure_50m:hPa	float32			float	
prob_rime:p	float32			float	
rain_water:kgm2	float32			float	
relative_humidity_1000hPa:p	float32			float	
sfc_pressure:hPa	float32			float	
snow_depth:cm	float32			float	
snow_melt_10min:mm	float32			float	
snow_water:kgm2	float32			float	
sun_azimuth:d	float32			float	

sun_elevation:d	float32	float
super_cooled_liquid_water:kgm2	float32	float
t_1000hPa:K	float32	float
total_cloud_cover:p	float32	float
visibility:m	float32	float
weekday	int64	int
wind_speed_10m:ms	float32	float
wind_speed_u_10m:ms	float32	float
wind_speed_v_10m:ms	float32	float
wind_speed_w_1000hPa:ms	float32	float
year	int64	int

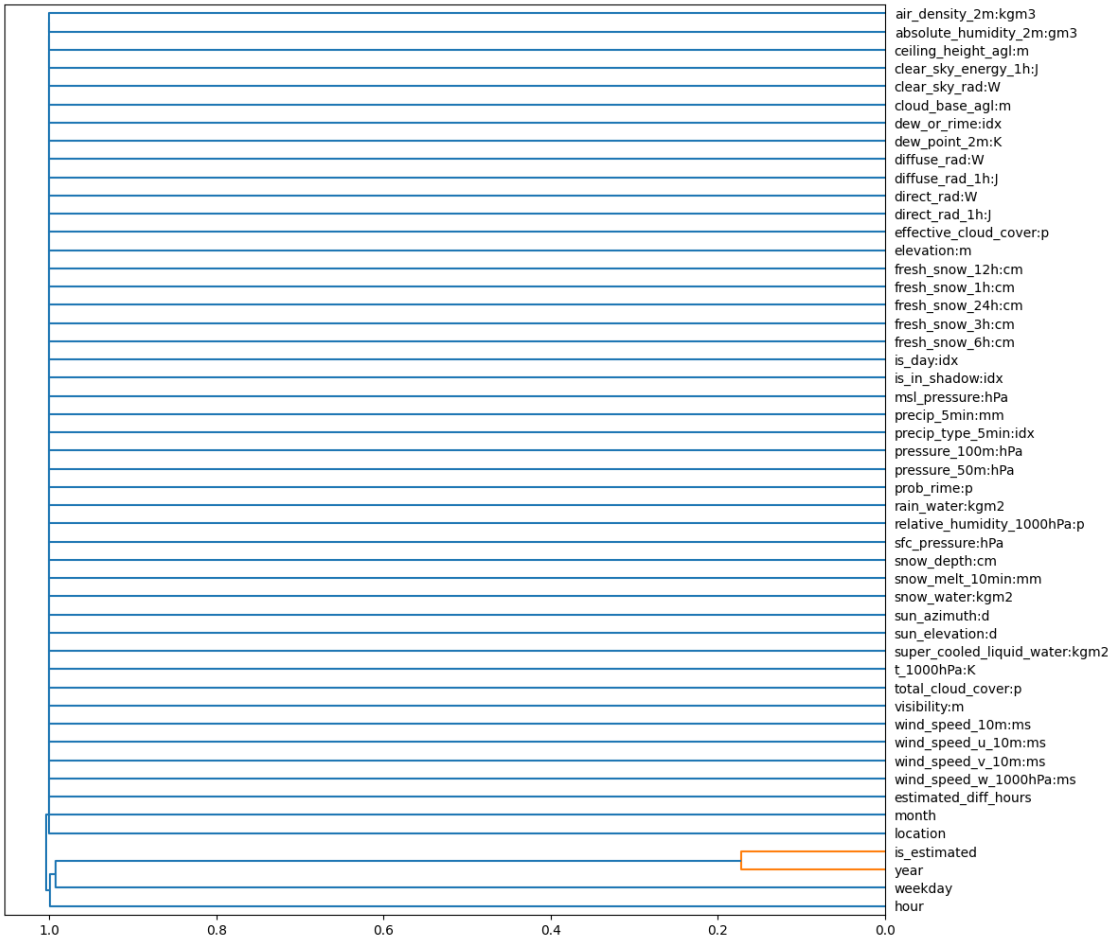
	variable_type	special_types
absolute_humidity_2m:gm3	numeric	
air_density_2m:kgm3	numeric	
ceiling_height_agl:m	numeric	
clear_sky_energy_1h:J	numeric	
clear_sky_rad:W	numeric	
cloud_base_agl:m	numeric	
dew_or_rime:idx	category	
dew_point_2m:K	numeric	
diffuse_rad:W	numeric	
diffuse_rad_1h:J	numeric	
direct_rad:W	numeric	
direct_rad_1h:J	numeric	
effective_cloud_cover:p	numeric	
elevation:m	category	
estimated_diff_hours	numeric	
fresh_snow_12h:cm	category	
fresh_snow_1h:cm	category	
fresh_snow_24h:cm	category	
fresh_snow_3h:cm	category	
fresh_snow_6h:cm	category	
hour	numeric	
is_day:idx	category	
is_estimated	category	
is_in_shadow:idx	category	
location	category	
month	category	
msl_pressure:hPa	numeric	
precip_5min:mm	numeric	
precip_type_5min:idx	category	
pressure_100m:hPa	numeric	
pressure_50m:hPa	numeric	
prob_rime:p	category	
rain_water:kgm2	category	
relative_humidity_1000hPa:p	numeric	
sfc_pressure:hPa	numeric	

snow_depth:cm	category
snow_melt_10min:mm	category
snow_water:kgm2	category
sun_azimuth:d	numeric
sun_elevation:d	numeric
super_cooled_liquid_water:kgm2	category
t_1000hPa:K	numeric
total_cloud_cover:p	numeric
visibility:m	numeric
weekday	category
wind_speed_10m:ms	numeric
wind_speed_u_10m:ms	numeric
wind_speed_v_10m:ms	numeric
wind_speed_w_1000hPa:ms	category
year	category

Types warnings summary

	train_data	test_data	warnings
estimated_diff_hours	float	int	warning
is_estimated	object	bool	warning
y	float	--	warning

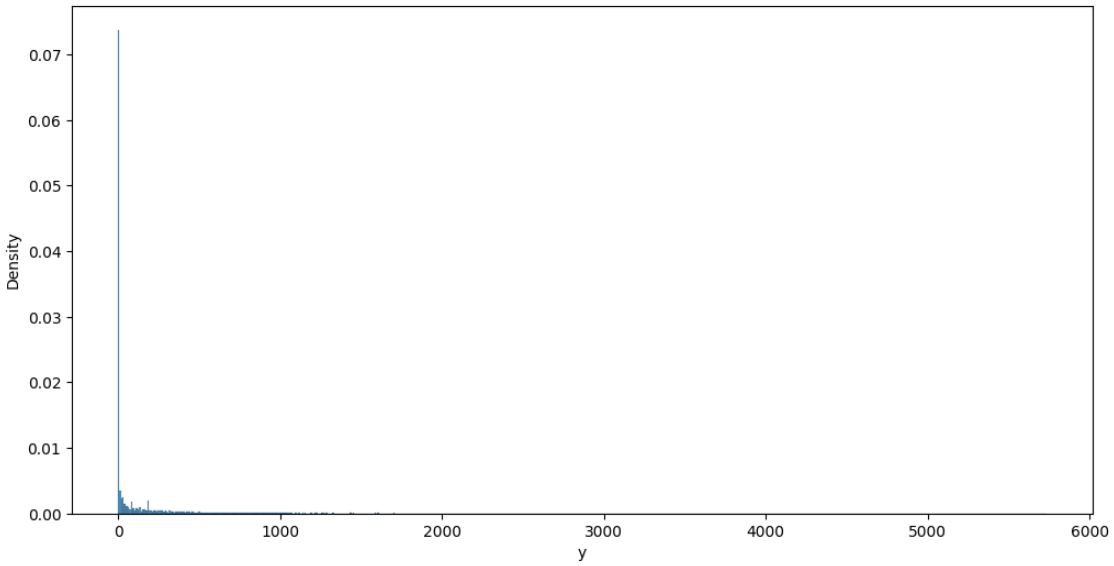
1.0.1 Feature Distance



[5]:
```
auto.target_analysis(train_data=X_train, label="y")#, sample=None)
```

1.1 Target variable analysis

```
   count        mean         std  min  25%  50%     75%      max   dtypes  \
y  93024  287.019652  766.407785 -0.0  0.0  0.0  172.92  5733.42  float64

   unique missing_count missing_ratio raw_type special_types
y   12430                                float
```

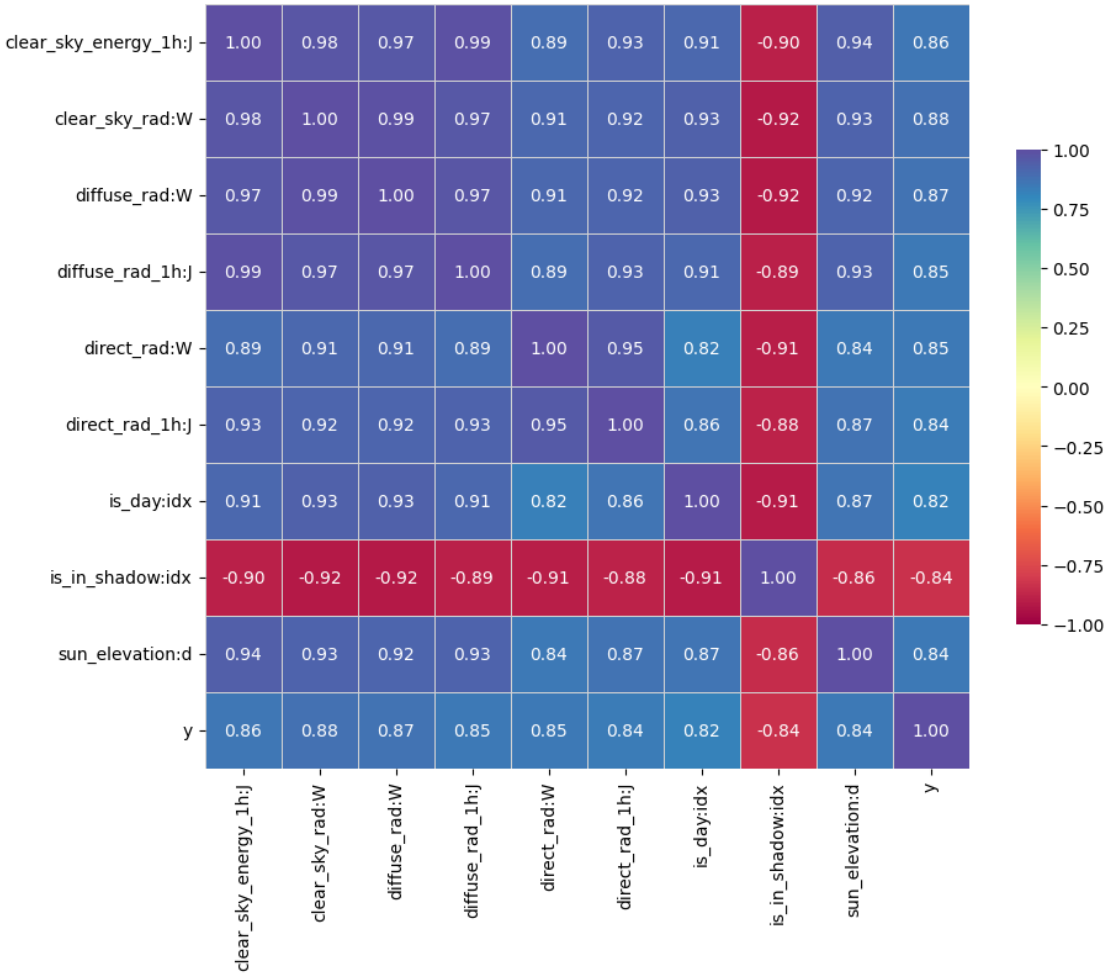


1.1.1 Distribution fits for target variable

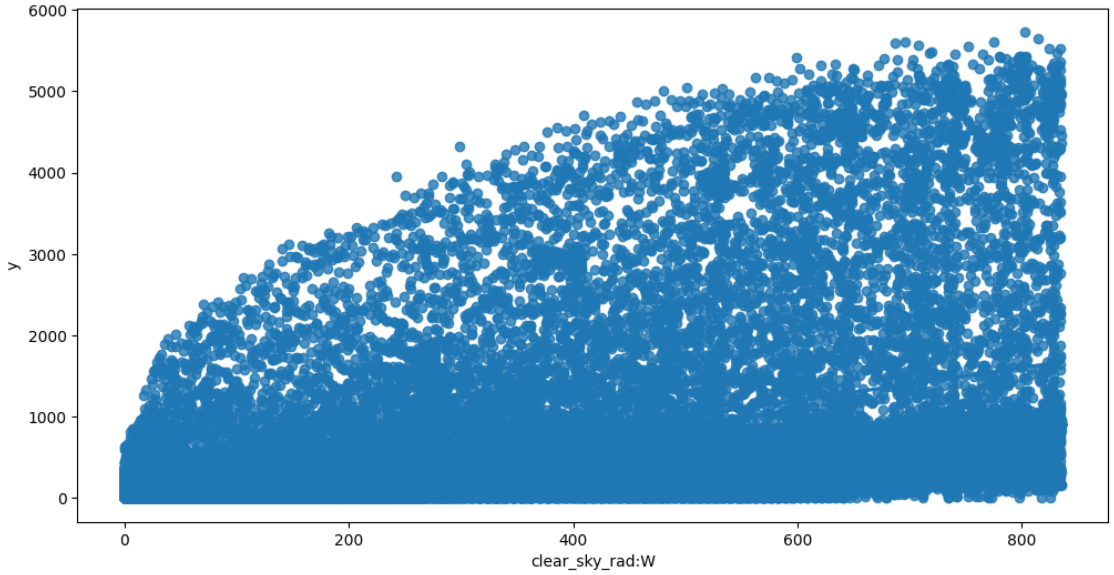
- none of the attempted distribution fits satisfy specified minimum p-value threshold: `0.01`

1.1.2 Target variable correlations

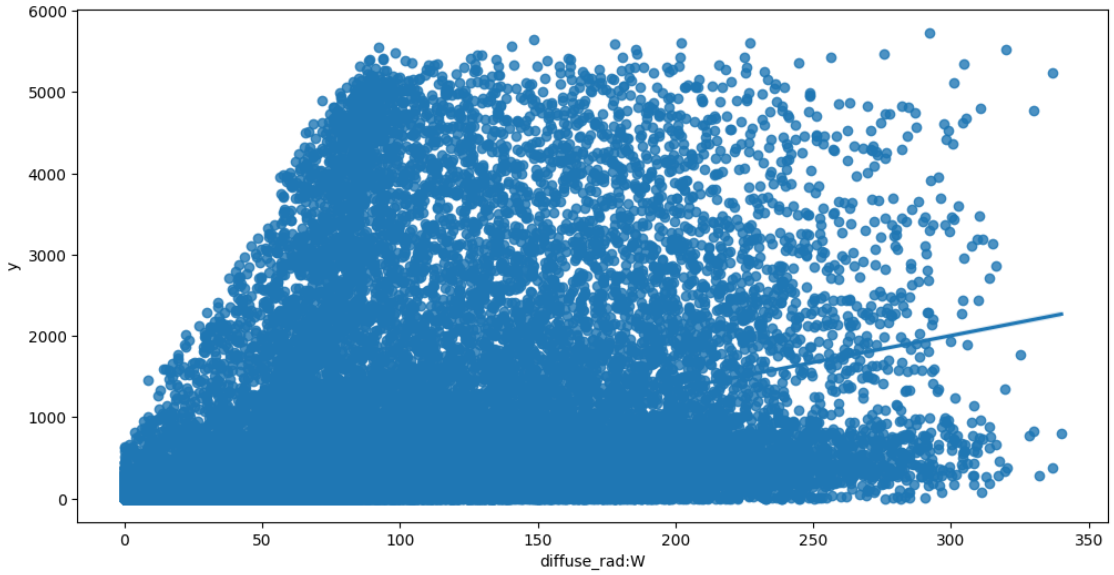
`train_data` - `spearman` correlation matrix; focus: absolute correlation for `y` >= `0.5`



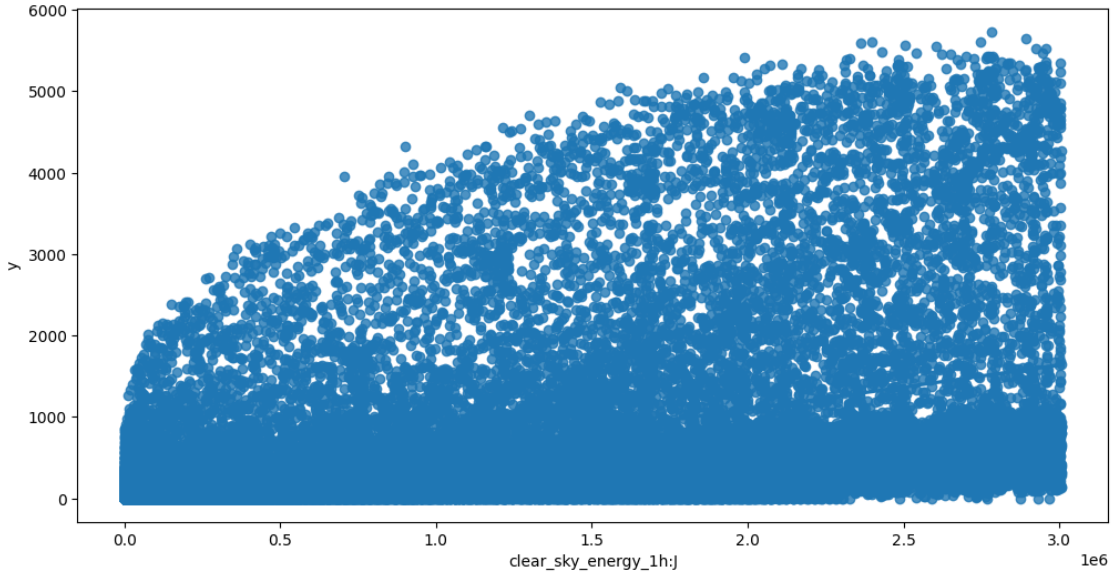
Feature interaction between `clear_sky_rad:W`/`y` in `train_data`



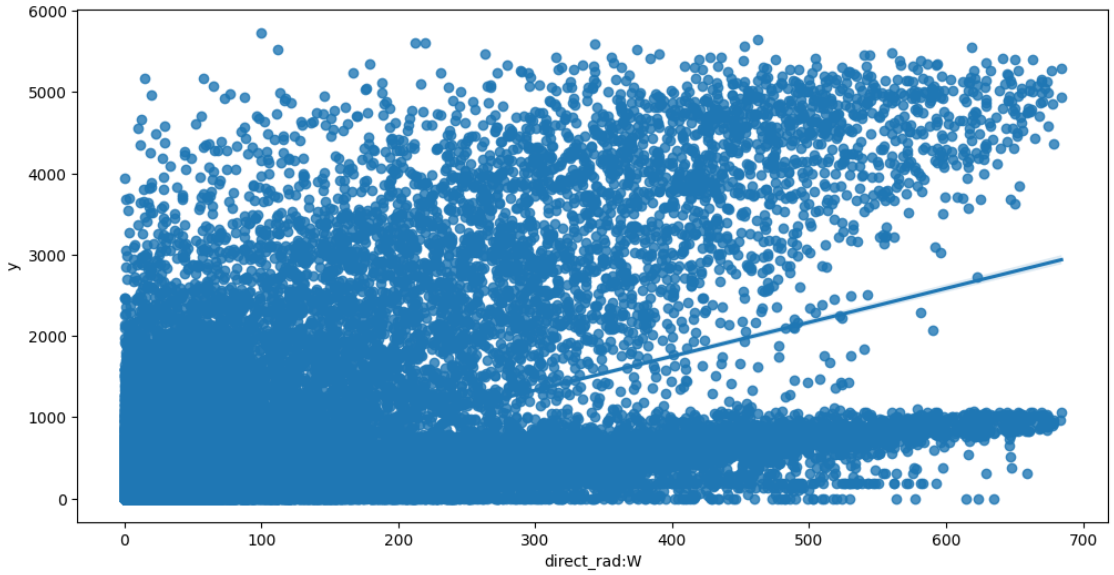
Feature interaction between `diffuse_rad:W/y` in `train_data`



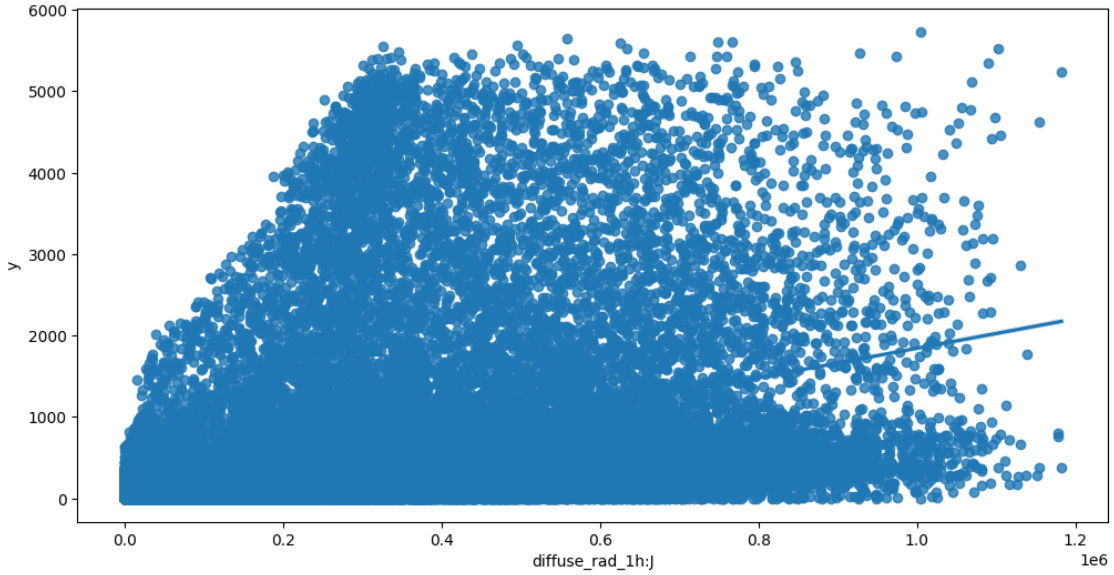
Feature interaction between `clear_sky_energy_1h:J/y` in `train_data`



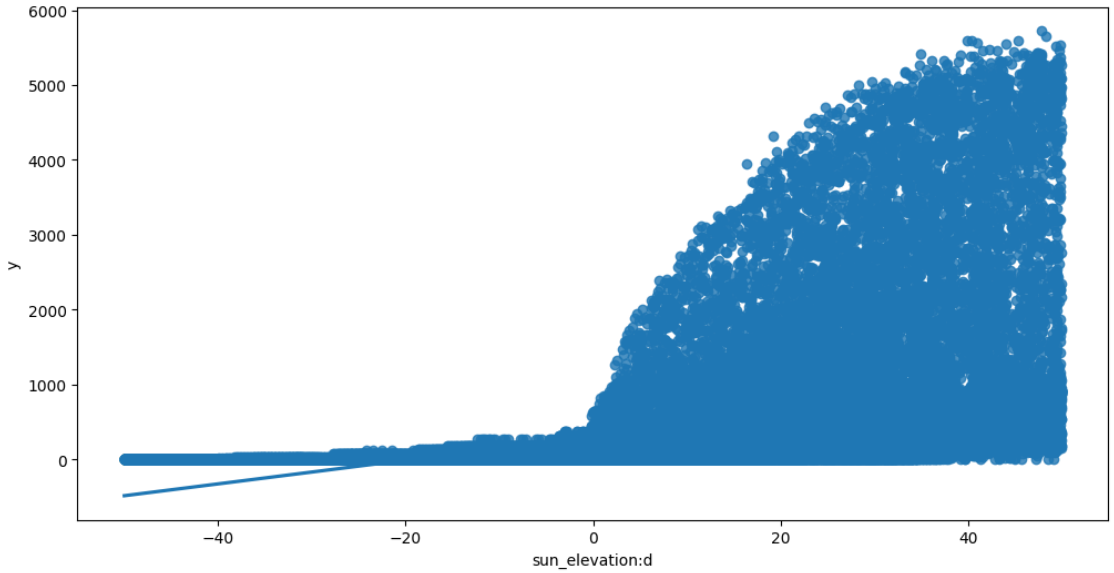
Feature interaction between `direct_rad:W/y` in `train_data`



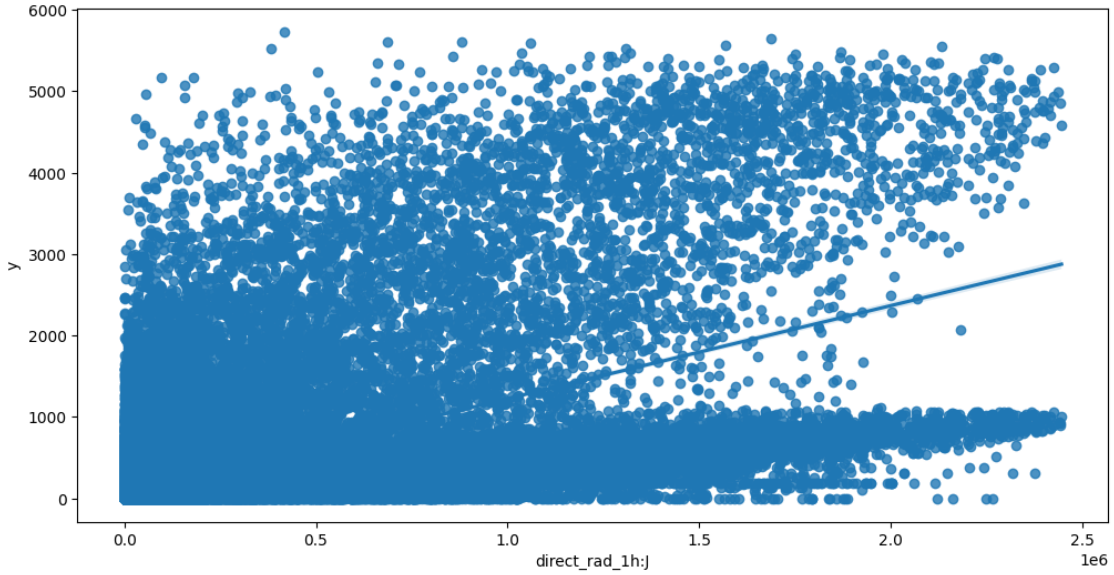
Feature interaction between `diffuse_rad_1h:J/y` in `train_data`



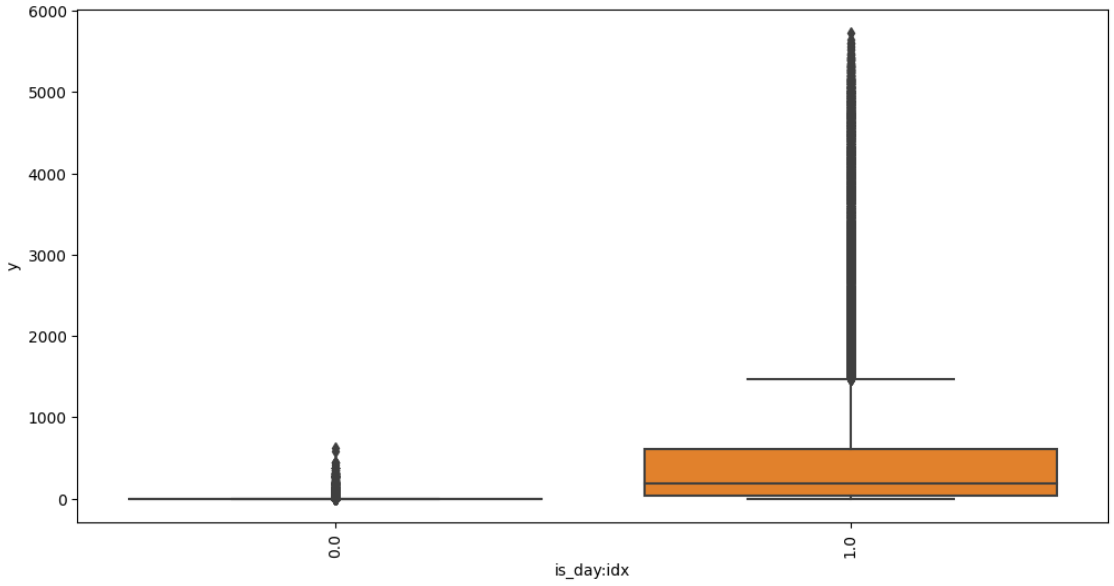
Feature interaction between `sun_elevation:d/y` in `train_data`



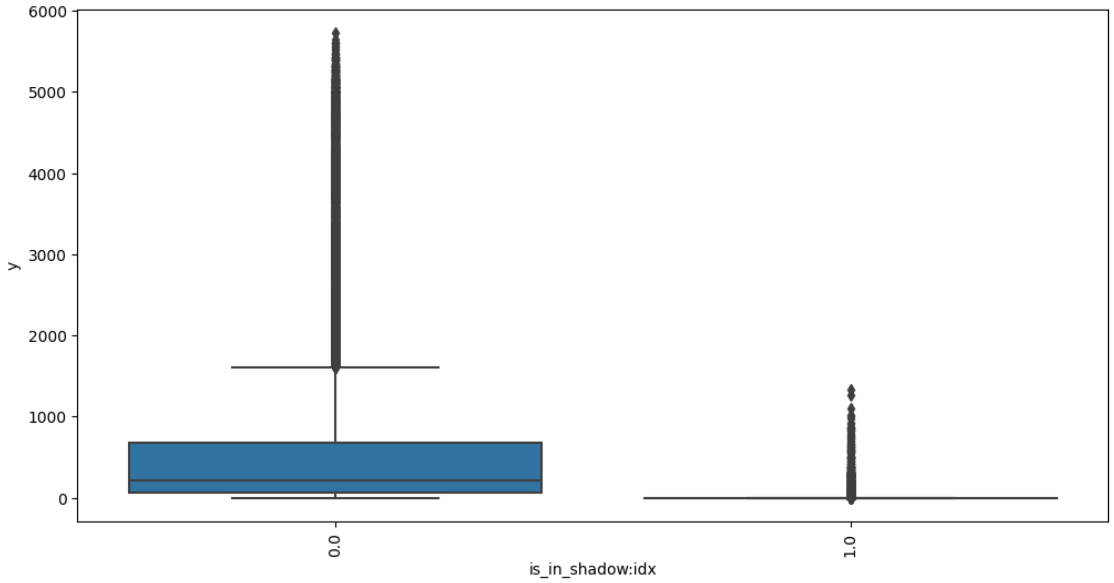
Feature interaction between `direct_rad_1h:J/y` in `train_data`



Feature interaction between `is_day:idx/y` in `train_data`



Feature interaction between `is_in_shadow:idx/y` in `train_data`

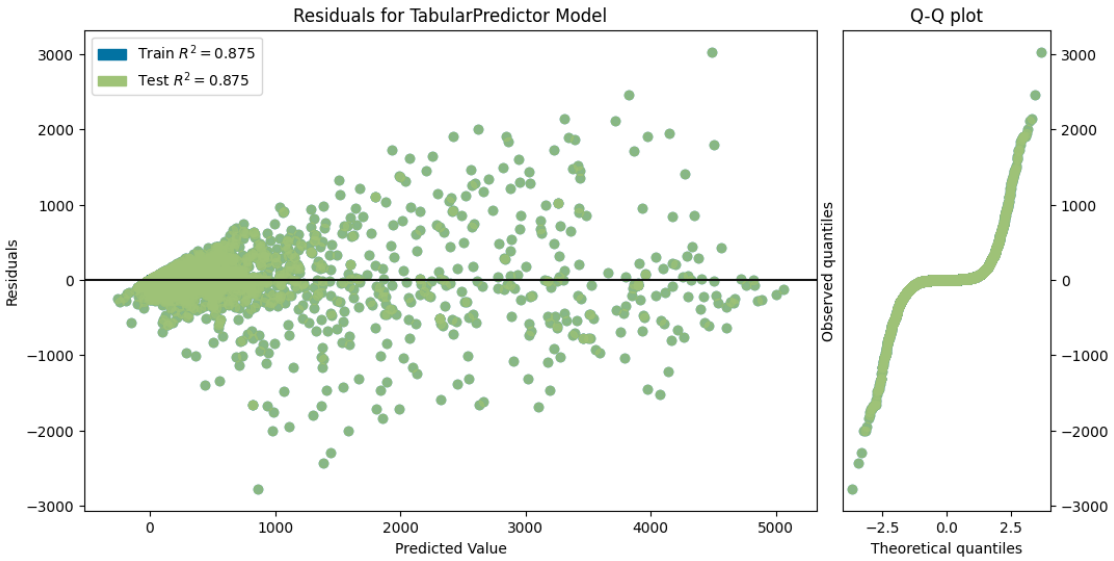


[14]:
```
auto.quick_fit(X_train, "y", show_feature_importance_barplots=True, val_size=0.
↪3, sample=20000)
```

```
No path specified. Models will be saved in:
"AutogluonModels/ag-20231007_072957/"
```

1.1.3 Model Prediction for y

Using validation data for `Test` points



1.1.4 Model Leaderboard

```
        model  score_test   score_val  pred_time_test  pred_time_val  \
0  LightGBMXT -248.750281 -219.751418        0.395301       0.089271

    fit_time  pred_time_test_marginal  pred_time_val_marginal  \
0  27.446321                0.395301                0.089271

   fit_time_marginal  stack_level  can_infer  fit_order
0          27.446321            1       True          1
```

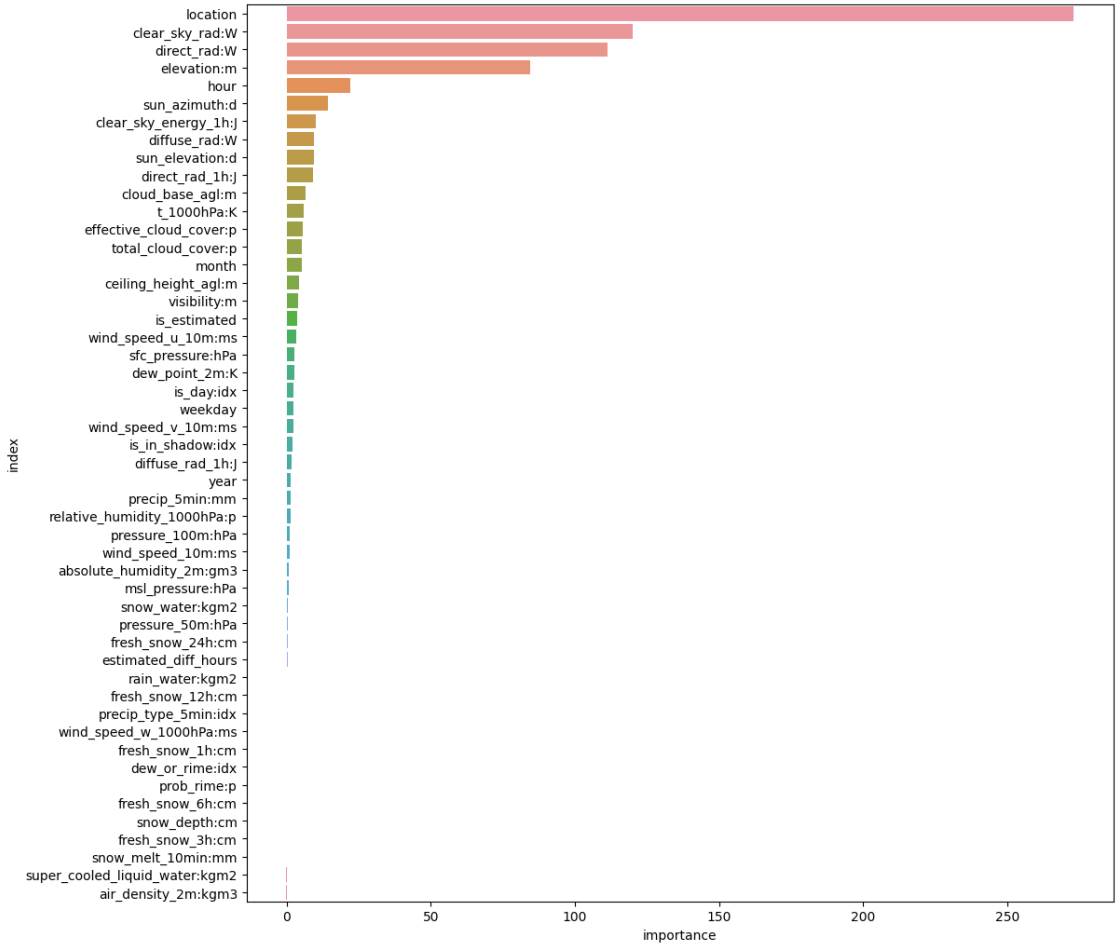
1.1.5 Feature Importance for Trained Model

	importance	stddev	p_value	n
location	273.036713	13.376774	6.891504e-07	5
clear_sky_rad:W	120.016183	7.180915	1.530633e-06	5
direct_rad:W	111.426500	4.257681	2.553141e-07	5
elevation:m	84.471386	7.655542	8.007653e-06	5
hour	22.000719	1.605756	3.381223e-06	5
sun_azimuth:d	14.428784	1.146843	4.749289e-06	5
clear_sky_energy_1h:J	10.300921	2.154345	2.167830e-04	5
diffuse_rad:W	9.496224	1.404482	5.578025e-05	5
sun_elevation:d	9.463388	1.478064	6.914676e-05	5
direct_rad_1h:J	9.028461	1.675240	1.359428e-04	5
cloud_base_agl:m	6.653806	1.043902	7.037507e-05	5
t_1000hPa:K	5.806126	1.781938	9.430619e-04	5
effective_cloud_cover:p	5.701147	1.884051	1.244427e-03	5
total_cloud_cover:p	5.427388	2.299403	3.089474e-03	5
month	5.322655	1.650851	9.811879e-04	5
ceiling_height_agl:m	4.366356	1.499439	1.435626e-03	5
visibility:m	4.041856	0.925590	3.081491e-04	5
is_estimated	3.776888	0.947758	4.383524e-04	5
wind_speed_u_10m:ms	3.305182	0.908822	6.219493e-04	5
sfc_pressure:hPa	2.673507	0.730667	6.076803e-04	5
dew_point_2m:K	2.619011	1.427539	7.412175e-03	5
is_day:idx	2.412707	0.269573	1.839403e-05	5
weekday	2.266859	2.301478	4.620089e-02	5
wind_speed_v_10m:ms	2.254496	1.695459	2.050383e-02	5
is_in_shadow:idx	2.059541	0.385591	1.407923e-04	5
diffuse_rad_1h:J	1.870572	0.903431	4.904257e-03	5
year	1.516330	1.139311	2.044974e-02	5
precip_5min:mm	1.488578	0.908988	1.077274e-02	5
relative_humidity_1000hPa:p	1.338183	0.817485	1.078699e-02	5
pressure_100m:hPa	1.236986	0.606941	5.180194e-03	5
wind_speed_10m:ms	1.210337	1.074450	3.271787e-02	5
absolute_humidity_2m:gm3	0.752585	0.890510	6.589527e-02	5
msl_pressure:hPa	0.743825	0.505388	1.509325e-02	5
snow_water:kgm2	0.516878	0.599791	6.312787e-02	5

pressure_50m:hPa	0.465997	0.874107	1.495594e-01	5
fresh_snow_24h:cm	0.304835	0.270573	3.270585e-02	5
estimated_diff_hours	0.289331	0.328694	6.020290e-02	5
rain_water:kgm2	0.123238	0.206141	1.261270e-01	5
fresh_snow_12h:cm	0.046842	0.234500	3.391171e-01	5
precip_type_5min:idx	0.037837	0.280645	3.890393e-01	5
wind_speed_w_1000hPa:ms	0.005803	0.005487	3.861826e-02	5
fresh_snow_1h:cm	0.001067	0.006419	3.645258e-01	5
dew_or_rime:idx	-0.000973	0.064345	5.126739e-01	5
prob_rime:p	-0.003272	0.005454	8.745846e-01	5
fresh_snow_6h:cm	-0.006316	0.120775	5.437299e-01	5
snow_depth:cm	-0.006965	0.094136	5.616913e-01	5
fresh_snow_3h:cm	-0.013236	0.027432	8.293228e-01	5
snow_melt_10min:mm	-0.030548	0.047669	8.874220e-01	5
super_cooled_liquid_water:kgm2	-0.122223	0.292552	7.984540e-01	5
air_density_2m:kgm3	-0.225525	0.333134	8.976786e-01	5

	p99_high	p99_low
location	300.579676	245.493751
clear_sky_rad:W	134.801787	105.230580
direct_rad:W	120.193124	102.659875
elevation:m	100.234252	68.708520
hour	25.306993	18.694445
sun_azimuth:d	16.790149	12.067418
clear_sky_energy_1h:J	14.736747	5.865094
diffuse_rad:W	12.388071	6.604377
sun_elevation:d	12.506742	6.420034
direct_rad_1h:J	12.477803	5.579119
cloud_base_agl:m	8.803214	4.504398
t_1000hPa:K	9.475160	2.137092
effective_cloud_cover:p	9.580434	1.821860
total_cloud_cover:p	10.161890	0.692886
month	8.721779	1.923530
ceiling_height_agl:m	7.453723	1.278990
visibility:m	5.947659	2.136052
is_estimated	5.728334	1.825442
wind_speed_u_10m:ms	5.176460	1.433904
sfc_pressure:hPa	4.177961	1.169053
dew_point_2m:K	5.558335	-0.320312
is_day:idx	2.967762	1.857651
weekday	7.005634	-2.471917
wind_speed_v_10m:ms	5.745470	-1.236478
is_in_shadow:idx	2.853479	1.265603
diffuse_rad_1h:J	3.730748	0.010395
year	3.862186	-0.829526
precip_5min:mm	3.360197	-0.383040
relative_humidity_1000hPa:p	3.021397	-0.345031
pressure_100m:hPa	2.486686	-0.012714

wind_speed_10m:ms	3.422644	-1.001970
absolute_humidity_2m:gm3	2.586158	-1.080988
msl_pressure:hPa	1.784426	-0.296776
snow_water:kgm2	1.751856	-0.718099
pressure_50m:hPa	2.265796	-1.333801
fresh_snow_24h:cm	0.861949	-0.252279
estimated_diff_hours	0.966116	-0.387454
rain_water:kgm2	0.547687	-0.301210
fresh_snow_12h:cm	0.529681	-0.435996
precip_type_5min:idx	0.615689	-0.540015
wind_speed_w_1000hPa:ms	0.017101	-0.005494
fresh_snow_1h:cm	0.014284	-0.012151
dew_or_rime:idx	0.131514	-0.133460
prob_rime:p	0.007957	-0.014502
fresh_snow_6h:cm	0.242361	-0.254994
snow_depth:cm	0.186862	-0.200793
fresh_snow_3h:cm	0.043248	-0.069720
snow_melt_10min:mm	0.067602	-0.128698
super_cooled_liquid_water:kgm2	0.480145	-0.724592
air_density_2m:kgm3	0.460402	-0.911451



1.1.6 Rows with the highest prediction error

Rows in this category worth inspecting for the causes of the error

ds	absolute_humidity_2m:gm3	air_density_2m:kgm3	\
2021-07-08 11:00:00	11.1	1.185	
2019-09-01 11:00:00	11.1	1.206	
2023-04-16 08:00:00	7.0	1.264	
2022-08-26 08:00:00	12.4	1.200	
2021-10-05 09:00:00	8.5	1.218	
2019-07-09 11:00:00	9.7	1.221	
2021-06-19 07:00:00	7.3	1.229	
2022-06-16 10:00:00	9.9	1.228	
2021-07-06 14:00:00	13.4	1.172	
2022-05-24 12:00:00	6.2	1.186	

ds	ceiling_height_agl:m	clear_sky_energy_1h:J	\
2021-07-08 11:00:00	8978.200195	2896551.500	
2019-09-01 11:00:00	1206.400024	2103079.250	
2023-04-16 08:00:00	NaN	1344241.625	
2022-08-26 08:00:00	889.099976	1338429.625	
2021-10-05 09:00:00	6272.600098	801049.125	
2019-07-09 11:00:00	1523.800049	2897436.250	
2021-06-19 07:00:00	9630.900391	1675875.500	
2022-06-16 10:00:00	2031.800049	2788318.750	
2021-07-06 14:00:00	3621.699951	2683876.500	
2022-05-24 12:00:00	6346.500000	2879051.500	

ds	clear_sky_rad:W	cloud_base_agl:m	dew_or_rime:idx	\
2021-07-08 11:00:00	818.299988	1731.400024	0.0	
2019-09-01 11:00:00	596.599976	1206.400024	0.0	
2023-04-16 08:00:00	432.799988	131.699997	0.0	
2022-08-26 08:00:00	430.700012	620.500000	0.0	
2021-10-05 09:00:00	264.399994	3214.000000	0.0	
2019-07-09 11:00:00	817.900024	1523.800049	0.0	
2021-06-19 07:00:00	529.099976	1865.800049	0.0	
2022-06-16 10:00:00	803.799988	698.500000	0.0	
2021-07-06 14:00:00	706.799988	3180.000000	0.0	
2022-05-24 12:00:00	792.299988	2348.600098	0.0	

ds	dew_point_2m:K	diffuse_rad:W	diffuse_rad_1h:J	…	\
				…	

```
2021-07-08 11:00:00      286.399994     148.199997      492451.31250  …
2019-09-01 11:00:00      285.899994     130.100006      346342.40625  …
2023-04-16 08:00:00      278.700012     105.199997      364302.18750  …
2022-08-26 08:00:00      287.799988     126.900002      359059.40625  …
2021-10-05 09:00:00      281.799988     107.599998      324482.09375  …
2019-07-09 11:00:00      284.000000     226.399994      802221.18750  …
2021-06-19 07:00:00      279.700012     115.800003      387195.90625  …
2022-06-16 10:00:00      284.200012     221.199997      754735.12500  …
2021-07-06 14:00:00      289.299988     212.100006      682230.87500  …
2022-05-24 12:00:00      277.600006     238.199997      899213.68750  …

                     estimated_diff_hours  is_estimated  location  hour  \
ds
2021-07-08 11:00:00                   0.0         False         A    11
2019-09-01 11:00:00                   0.0         False         A    11
2023-04-16 08:00:00                  24.0          True         A     8
2022-08-26 08:00:00                   0.0         False         A     8
2021-10-05 09:00:00                   0.0         False         A     9
2019-07-09 11:00:00                   0.0         False         A    11
2021-06-19 07:00:00                   0.0         False         A     7
2022-06-16 10:00:00                   0.0         False         A    10
2021-07-06 14:00:00                   0.0         False         A    14
2022-05-24 12:00:00                   0.0         False         A    12

                     weekday  month  year        y       y_pred        error
ds
2021-07-08 11:00:00        3      7  2021  1450.02  4482.781250  3032.761250
2019-09-01 11:00:00        6      9  2019  3643.20   865.851746  2777.348254
2023-04-16 08:00:00        6      4  2023  1364.44  3820.209473  2455.769473
2022-08-26 08:00:00        4      8  2022  3821.84  1384.216309  2437.623691
2021-10-05 09:00:00        1     10  2021  3742.42  1439.572144  2302.847856
2019-07-09 11:00:00        1      7  2019  1172.60  3311.417969  2138.817969
2021-06-19 07:00:00        5      6  2021  1591.70  3712.722656  2121.022656
2022-06-16 10:00:00        3      6  2022  3589.74  1582.717529  2007.022471
2021-07-06 14:00:00        1      7  2021  2983.64   977.366455  2006.273545
2022-05-24 12:00:00        1      5  2022   621.72  2620.532715  1998.812715

[10 rows x 53 columns]
```

2 Starting

[15]:
```python
import os


# Get the last submission number
last_submission_number = int(max([int(filename.split('_')[1].split('.')[0]) for␣
 ↪filename in os.listdir('submissions') if "submission" in filename]))
```

```
print("Last submission number:", last_submission_number)
print("Now creating submission number:", last_submission_number + 1)

# Create the new filename
new_filename = f'submission_{last_submission_number + 1}'

hello = os.environ.get('HELLO')
if hello is not None:
    new_filename += f'_{hello}'

print("New filename:", new_filename)
```

```
Last submission number: 81
Now creating submission number: 82
New filename: submission_82_jorge
```

[16]:
```
from autogluon.tabular import TabularDataset, TabularPredictor
train_data = TabularDataset('X_train_raw.csv')
train_data.drop(columns=['ds'], inplace=True)

label = 'y'
metric = 'mean_absolute_error'
time_limit = 60*3
presets = 'best_quality'
```

```
Loaded data from: X_train_raw.csv | Columns = 52 / 52 | Rows = 136724 -> 136724
```

[17]:
```
predictor = TabularPredictor(label=label, eval_metric=metric,␣
  ↪path=f"AutogluonModels/{new_filename}").fit(train_data, presets=presets,␣
  ↪time_limit=time_limit)
```

```
Presets specified: ['best_quality']
Stack configuration (auto_stack=True): num_stack_levels=1, num_bag_folds=8,
num_bag_sets=20
Beginning AutoGluon training … Time limit = 180s
AutoGluon will save models to "AutogluonModels/submission_82_jorge/"
AutoGluon Version:  0.8.1
Python Version:     3.10.12
Operating System:   Darwin
Platform Machine:   arm64
Platform Version:   Darwin Kernel Version 22.1.0: Sun Oct  9 20:15:09 PDT 2022;
root:xnu-8792.41.9~2/RELEASE_ARM64_T6000
Disk Space Avail:   19.68 GB / 494.38 GB (4.0%)
Train Data Rows:    136724
Train Data Columns: 50
Label Column: y
Preprocessing data …
AutoGluon infers your prediction problem is: 'regression' (because dtype of
label-column == float and many unique label-values observed).
```

```
        Label info (max, min, mean, stddev): (5733.42, -0.0, 247.8577,
717.45424)
        If 'regression' is not the correct problem_type, please manually specify
the problem_type parameter during predictor init (You may specify problem_type
as one of: ['binary', 'multiclass', 'regression'])
Using Feature Generators to preprocess the data …
Fitting AutoMLPipelineFeatureGenerator…
        Available Memory:                    6004.48 MB
        Train Data (Original)  Memory Usage: 64.81 MB (1.1% of available memory)
        Inferring data type of each feature based on column values. Set
feature_metadata_in to manually specify special dtypes of the features.
        Stage 1 Generators:
                Fitting AsTypeFeatureGenerator…
        Stage 2 Generators:
                Fitting FillNaFeatureGenerator…
        Stage 3 Generators:
                Fitting IdentityFeatureGenerator…
                Fitting CategoryFeatureGenerator…
                        Fitting CategoryMemoryMinimizeFeatureGenerator…
        Stage 4 Generators:
                Fitting DropUniqueFeatureGenerator…
        Stage 5 Generators:
                Fitting DropDuplicatesFeatureGenerator…
        Types of features in original data (raw dtype, special dtypes):
                ('float', [])  : 44 | ['absolute_humidity_2m:gm3',
'air_density_2m:kgm3', 'ceiling_height_agl:m', 'clear_sky_energy_1h:J',
'clear_sky_rad:W', …]
                ('int', [])    :  4 | ['hour', 'weekday', 'month', 'year']
                ('object', []) :  2 | ['is_estimated', 'location']
        Types of features in processed data (raw dtype, special dtypes):
                ('category', []) :  2 | ['is_estimated', 'location']
                ('float', [])    : 44 | ['absolute_humidity_2m:gm3',
'air_density_2m:kgm3', 'ceiling_height_agl:m', 'clear_sky_energy_1h:J',
'clear_sky_rad:W', …]
                ('int', [])      :  4 | ['hour', 'weekday', 'month', 'year']
        0.4s = Fit runtime
        50 features in original data used to generate 50 features in processed
data.
        Train Data (Processed) Memory Usage: 52.78 MB (0.9% of available memory)
Data preprocessing and feature engineering runtime = 0.46s …
AutoGluon will gauge predictive performance using evaluation metric:
'mean_absolute_error'
        This metric's sign has been flipped to adhere to being higher_is_better.
The metric score can be multiplied by -1 to get the metric value.
        To change this, specify the eval_metric parameter of Predictor()
User-specified model hyperparameters to be fit:
{
        'NN_TORCH': {},
```

```
        'GBM': [{'extra_trees': True, 'ag_args': {'name_suffix': 'XT'}}, {},
'GBMLarge'],
        'CAT': {},
        'XGB': {},
        'FASTAI': {},
        'RF': [{'criterion': 'gini', 'ag_args': {'name_suffix': 'Gini',
'problem_types': ['binary', 'multiclass']}}, {'criterion': 'entropy', 'ag_args':
{'name_suffix': 'Entr', 'problem_types': ['binary', 'multiclass']}},
{'criterion': 'squared_error', 'ag_args': {'name_suffix': 'MSE',
'problem_types': ['regression', 'quantile']}}],
        'XT': [{'criterion': 'gini', 'ag_args': {'name_suffix': 'Gini',
'problem_types': ['binary', 'multiclass']}}, {'criterion': 'entropy', 'ag_args':
{'name_suffix': 'Entr', 'problem_types': ['binary', 'multiclass']}},
{'criterion': 'squared_error', 'ag_args': {'name_suffix': 'MSE',
'problem_types': ['regression', 'quantile']}}],
        'KNN': [{'weights': 'uniform', 'ag_args': {'name_suffix': 'Unif'}},
{'weights': 'distance', 'ag_args': {'name_suffix': 'Dist'}}],
}
AutoGluon will fit 2 stack levels (L1 to L2) …
Fitting 11 L1 models …
Fitting model: KNeighborsUnif_BAG_L1 … Training model for up to 119.67s of the
179.54s of remaining time.
        Not enough time to generate out-of-fold predictions for model. Estimated
time required was 2282.19s compared to 155.45s of available time.
        Time limit exceeded… Skipping KNeighborsUnif_BAG_L1.
Fitting model: KNeighborsDist_BAG_L1 … Training model for up to 111.16s of the
171.03s of remaining time.
        Not enough time to generate out-of-fold predictions for model. Estimated
time required was 1914.93s compared to 144.4s of available time.
        Time limit exceeded… Skipping KNeighborsDist_BAG_L1.
Fitting model: LightGBMXT_BAG_L1 … Training model for up to 104.0s of the
163.88s of remaining time.
        Fitting 8 child models (S1F1 - S1F8) | Fitting with
ParallelLocalFoldFittingStrategy
        Warning: Exception caused LightGBMXT_BAG_L1 to fail during training…
Skipping this model.
                [Errno 1] Operation not permitted
Detailed Traceback:
Traceback (most recent call last):
  File "/opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-
packages/autogluon/core/trainer/abstract_trainer.py", line 1733, in
_train_and_save
    model = self._train_single(X, y, model, X_val, y_val,
total_resources=total_resources, **model_fit_kwargs)
  File "/opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-
packages/autogluon/core/trainer/abstract_trainer.py", line 1684, in
_train_single
    model = model.fit(X=X, y=y, X_val=X_val, y_val=y_val,
```

```
total_resources=total_resources, **model_fit_kwargs)
  File "/opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-
packages/autogluon/core/models/abstract/abstract_model.py", line 829, in fit
    out = self._fit(**kwargs)
  File "/opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-
packages/autogluon/core/models/ensemble/stacker_ensemble_model.py", line 169, in
_fit
    return super()._fit(X=X, y=y, time_limit=time_limit, **kwargs)
  File "/opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-
packages/autogluon/core/models/ensemble/bagged_ensemble_model.py", line 266, in
_fit
    self._fit_folds(
  File "/opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-
packages/autogluon/core/models/ensemble/bagged_ensemble_model.py", line 592, in
_fit_folds
    fold_fitting_strategy.after_all_folds_scheduled()
  File "/opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-
packages/autogluon/core/models/ensemble/fold_fitting_strategy.py", line 508, in
after_all_folds_scheduled
    self.ray.init(**ray_init_args)
  File "/Users/jorgensandhaug/.local/lib/python3.10/site-
packages/ray/_private/client_mode_hook.py", line 105, in wrapper
    return func(*args, **kwargs)
  File "/Users/jorgensandhaug/.local/lib/python3.10/site-
packages/ray/_private/worker.py", line 1555, in init
    connect(
  File "/Users/jorgensandhaug/.local/lib/python3.10/site-
packages/ray/_private/worker.py", line 1926, in connect
    faulthandler.enable(all_threads=False)
  File "/opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-
packages/ipykernel/kernelapp.py", line 526, in enable
    return faulthandler_enable(file=file, all_threads=all_threads, **kwargs)
PermissionError: [Errno 1] Operation not permitted
Fitting model: LightGBM_BAG_L1 … Training model for up to 101.16s of the
161.04s of remaining time.
        Fitting 8 child models (S1F1 - S1F8) | Fitting with
ParallelLocalFoldFittingStrategy
```

```
The Kernel crashed while executing code in the the current cell or a previous␣
 ↪cell. Please review the code in the cell(s) to identify a possible cause of␣
 ↪the failure. Click <a href='https://aka.ms/vscodeJupyterKernelCrash'>here</a>␣
 ↪for more info. View Jupyter <a href='command:jupyter.viewOutput'>log</a> for␣
 ↪further details.
```

[ ]:
```
predictors = [predictor, predictor, predictor]
```

3 Submit

[ ]:
```
import pandas as pd
import matplotlib.pyplot as plt

train_data_with_dates = TabularDataset('X_train_raw.csv')
train_data_with_dates["ds"] = pd.to_datetime(train_data_with_dates["ds"])

test_data = TabularDataset('X_test_raw.csv')
test_data["ds"] = pd.to_datetime(test_data["ds"])
#test_data
```

[ ]:
```
test_ids = TabularDataset('test.csv')
test_ids["time"] = pd.to_datetime(test_ids["time"])
# merge test_data with test_ids
test_data_merged = pd.merge(test_data, test_ids, how="inner", right_on=["time",␣
 ↪"location"], left_on=["ds", "location"])

#test_data_merged
```

[ ]:
```
# predict, grouped by location
predictions = []
location_map = {
    "A": 0,
    "B": 1,
    "C": 2
}
for loc, group in test_data.groupby('location'):
    i = location_map[loc]
    subset = test_data_merged[test_data_merged["location"] == loc].
 ↪reset_index(drop=True)
    #print(subset)
    pred = predictors[i].predict(subset)
    subset["prediction"] = pred
    predictions.append(subset)
```

[ ]:
```
# plot predictions for location A, in addition to train data for A
for loc, idx in location_map.items():
    fig, ax = plt.subplots(figsize=(20, 10))
    # plot train data
    train_data_with_dates[train_data_with_dates["location"]==loc].plot(x='ds',␣
 ↪y='y', ax=ax, label="train data")

    # plot predictions
    predictions[idx].plot(x='ds', y='prediction', ax=ax, label="predictions")

    # title
```

```
    ax.set_title(f"Predictions for location {loc}")
```

[ ]:
```
# concatenate predictions
submissions_df = pd.concat(predictions)
submissions_df = submissions_df[["id", "prediction"]]
submissions_df
```

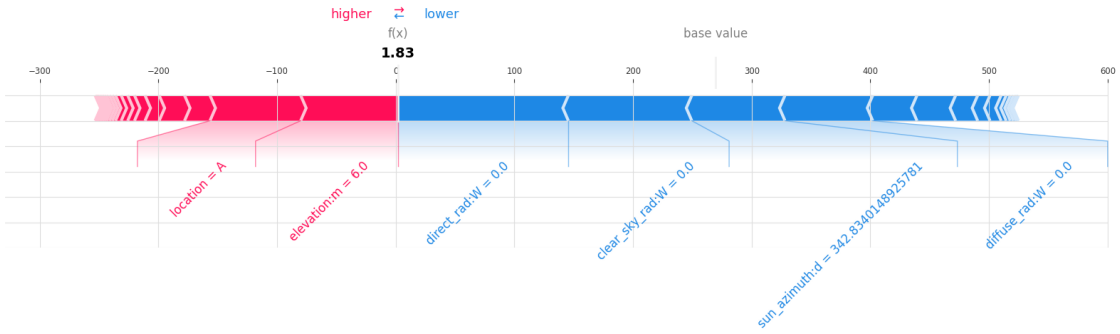
[ ]:
```
# Save the submission DataFrame to submissions folder, create new name based on␣
 ↪last submission, format is submission_<last_submission_number + 1>.csv

# Save the submission
print(f"Saving submission to submissions/{new_filename}.csv")
submissions_df.to_csv(os.path.join('submissions', f"{new_filename}.csv"),␣
 ↪index=False)
```

[ ]:
```
# save this notebook to submissions folder
import subprocess
import os
subprocess.run(["jupyter", "nbconvert", "--to", "pdf", "--output", os.path.
 ↪join('notebook_pdfs', f"{new_filename}.pdf"), "autogluon_all.ipynb"])
```

[ ]:
```
predictor.fit_summary(show_plot=True)
```

[ ]:
```
# feature importance
predictor.feature_importance(feature_stage="original",␣
 ↪data=train_data[train_data["location"] == "A"][-24*60*1:])
```

[ ]:
```
auto.explain_rows(train_data=X_train, model=predictor, plot="force",␣
 ↪rows=X_train[:2])
```



[ ]:
```
subprocess.run(["jupyter", "nbconvert", "--to", "pdf", "--output", os.path.
 ↪join('notebook_pdfs', f"{new_filename}_with_feature_importance.pdf"),␣
 ↪"autogluon_all.ipynb"])
```